MGMT 59000-148 Advanced Database Management

**Group Project - Team 4**

**Navigating the Skies**

Crafting a Data-Driven Entry Strategy for India's Aviation Sector

**Submitted by**

Akshita Sharma

Aravind Teja Chikoti

Pooja Udayanjali Kannuri

Rahul Chowdary Kunku

Sai Mona Duvvapu

Sathwik Kanukuntla

# A white letter with a black background Description automatically generatedGraduate students of

# Purdue University

# MS in Business Analytics & Information Management

# February 2023

**Table of Contents**

1.Background……...…………......………………………………..…..………………….02

2.Introduction……….………………………..………..……………………………...…..03

2.1 Objective…………….....…..…………………...……………..……………………03

3.Project Question………………………………………………………………………...03

4.Entity Relationship Diagram.………..………………………………..………………...04

5.Relational Data Model..………..…………………...……………..……………………05

6.Normalization...…...…………..…………………...……………..…………………..…05

7.Case Scenarios and Deliverables...………………..…………...… .……………….…...06

8.Conclusion...………..…………….………………..…………...… .……………….…...09

9.Appendix………….……….…………..….…...……………..……...………………......10

**1. Background**

The Indian aviation industry stands as one of the most rapidly expanding markets globally, driven by factors such as economic growth, urbanization, and an increase in disposable income among the middle class. The proliferation of low-cost carriers, coupled with government initiatives like the UDAN scheme aimed at enhancing regional connectivity, has democratized air travel, making it accessible to a wider demographic. Nonetheless, the sector is marked by fierce competition, with established players and ambitious new entrants vying for their share of the skies.

In such a vibrant yet saturated market, a data-driven entry strategy is not just prudent but essential for any new airline. The success of an entrant in the Indian aviation sector is predicated not only on competitive pricing but also on a deep understanding of the market dynamics, traveller profiles, and journey patterns. The essence of this strategy lies in leveraging granular data to derive actionable insights that inform every aspect of the business model, from pricing and route planning to customer service and marketing.

Entering the Indian aviation market requires a new airline to navigate through intense competition and varied traveller profiles with a data-driven strategy that prioritizes demographic insights and dynamic pricing. Initially, the focus should be on underserved segments and routes, where the airline can establish itself against market challengers by offering competitive fares and superior service quality. Understanding the demographics and city characteristics is crucial; for example, city with more literacy rates demand higher frequency and flexibility, while city with low literacy might prioritize cost-efficiency deals. By leveraging advanced SQL analytics to dissect historical booking data and market trends, the airline can develop a pricing model that dynamically adapts to fluctuating demand and traveller behaviour, ensuring optimal occupancy and profitability. This holistic strategy, which blends competitive pricing with a deep understanding of customer segmentation and route viability, will pave the way for the airline to scale its operations and eventually compete with market leaders.

**2. Introduction**

Our consultancy objective is to devise a strategic entry price point for an emergent airline in the highly competitive Indian aviation sector. We aim to harness the power of data to inform market-aligned pricing strategies. This objective will be realized through the meticulous processing and analysis of comprehensive flight, itinerary, and airport data sourced from a premier Indian travel application (sourced from Kaggle), augmented by the demographic variables of the origin and destination locations. Our focus will be to navigate the competitive landscape by initially targeting market challengers such as Air Asia, Akasa Air, Alliance Air, and Go First, thereby securing a stable market position before scaling to take on the established market leaders.

**2.1. Project Objectives:**

The overarching goal of our project is to deliver a data-driven, multifaceted pricing framework and acute market insights that cater to a diverse array of traveller profiles and journey patterns. Our deliverables encompass:

* A competitive pricing strategy that is attuned to temporal fluctuations, geographic considerations, and demographic diversity. This strategy will inform not only the pricing of tickets but also ancillary revenue streams, ensuring a competitive edge.
* An analysis of potential routes and itineraries grounded in historical data, aiding operational and logistics planning. This insight will support the airline in crafting a flight schedule that aligns with market demand and optimizes resource allocation.
* A comprehensive ranking analysis of Indian airports, providing a regional perspective on market entry points. This will enable the airline to prioritize its market entries, focusing on regions with the highest potential for growth and profitability.

Through these deliverables, we aspire to equip our client with a data-backed foundation for decision-making, positioning them effectively for a successful launch and sustained growth within the Indian aviation industry.

**3. Project Question**

How can a new airline in the Indian aviation sector establish a competitive entry price point and market strategy using advanced SQL analytics to process flight, itinerary, and airport data, combined with demographic insights of travel origins and destinations, to effectively compete with market challengers and eventually position itself against market leaders?

**4. Entity Relationship Diagram**

**A diagram of a flowchart

Description automatically generated**

Image 1: ERD

**Conceptual modelling rationale:**

The above Entity-Relationship Diagram (ERD) is a visual representation of the relationships between the **entities** as follows:

1. Airline
2. Travel\_Info
3. Airport
4. Place

**Description:**

* Airlines-Travel\_Info:
  + An Airline can have many Flights taking off.
  + A Flight taking off on given date and time can be associated with single airline.
* Travel\_Info-Airport:
  + A Flight can only take off from one airport on given date and time.
  + An Airport can be associated with multiple flights.
* Airport-Place:
  + An Airport can be associated with single Place.
  + A Place can have multiple airports

**5. Relational Data Model**

The image presents a clear and concise representation of the relational schema, complete with referential integrity constraints and functional dependencies.

* This schema serves as a foundational blueprint for organizing and structuring the database, ensuring that data relationships are maintained with precision and accuracy.
* The Referential integrity constraints guarantee the consistency and reliability of data by enforcing rules that govern the relationships between tables, while functional dependencies provide valuable insights into how attributes within the schema relate to one another.

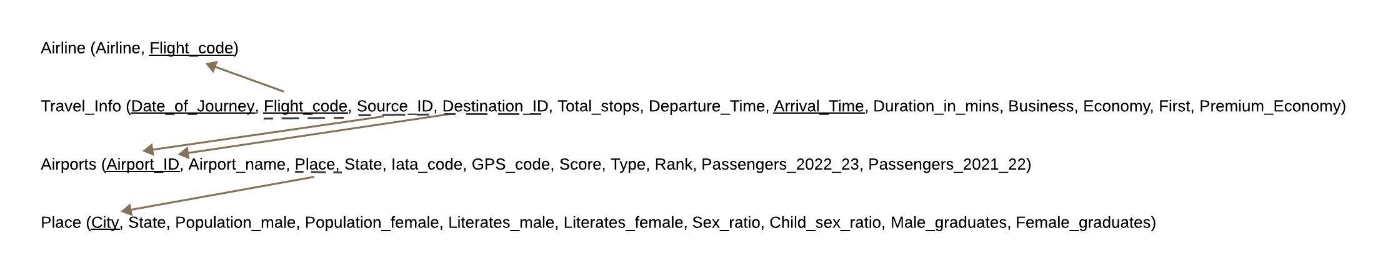


Image 2: Relational Schema

**6. Normalization**

Below is a snippet of the data on flights that was extracted from Kaggle.

**Raw data**



Image 3: Raw Data with flights information

There are some transitive functional dependencies that need to be eliminated to achieve 3NF. Furthermore, there are certain columns which convey the same information (Duration and duration in hours) that can be reduced.

Functional dependencies

1. **Date\_of\_journey** → Journey\_day
2. **Flight\_code** → Airline
3. **Departure\_time** → Departure
4. **Arrival\_time** → Arrival
5. **Departure\_time, Arrival\_time** → Duration
6. **Date\_of\_journey, Flight\_code, Arrival\_time** → Source, Destination, Total\_stops, Fare

By removing the columns with redundant information and by separating the airlines information to a new table, our final data set looks like the below where all the non-key attributes are dependent on all the three primary keys decided for the table. Furthermore, the fare information is now split to under four different columns based on class.



Image 4: Travel\_Info table in 3NF form.

Airline table has the flight\_codes along with the airlines they belong to. This is in 3NF form. Airports table has Airport\_ID as the sole primary key and all the attributes are dependent on this key, hence available in 3NF form. Place table has city as the primary key and all the other attributes are various metrics related specifically to each individual city. This is available in 3NF form, as well.

**6. Case Scenarios and Deliverables**

|  |  |  |  |
| --- | --- | --- | --- |
| **Topic** | **User Case Scenario** | **Client Deliverable** | **Key SQL techniques** |
| Economy class Pricing analysis | To determine the average price of economy class tickets for a specific competitor airline, broken down by source location | This analysis will help our client set competitive fares for their economy class seats in different regions. | * Procedures * Aggregations * Advanced Joins |
| Competitor fare analysis for different airport locations | To analyze and compare the average ticket prices of competitive airlines based on various factors: source place, source airport, destination place, and destination airport | This analysis will offer a panoramic view of the pricing landscape in the Indian aviation market | * Procedures * Advanced Joins |
| Ranking of destinations based on Departure Flight Volume from an airport | To identify and rank the top five most popular destination airports based on the number of flights departing for each | This analysis aims to highlight the busiest airports in terms of flight operations, providing insights into traffic concentration and operational scale | * Procedures * Advanced Joins |
| Hierarchical Analysis of Airport Popularity | To establish a hierarchy of airports based on their popularity | This analysis seeks to categorize airports into tiers, providing a structured perspective on their relative importance and influence in the aviation network | * Recursive CTE * Advanced Joins * Subqueries |
| Airport Ranking Based on Literacy Rates | To identify and rank the top 10 airports based on the literacy rates of males and females in their respective locations | This ranking will provide insights into the educational profile of the catchment areas for these airports, which can be crucial for targeted marketing and service design | * Subqueries |
| Airport Traffic analysis | To ascertain and rank the top 10 airports based on their passenger traffic | This analysis will shed light on the most frequented airports, which is indicative of high demand and potential profitability for airline routes | * Subqueries |
| Analysis of Airport Peak Departure Times | To identify the busiest departure times at various airports | This analysis aims to understand peak operational hours, which is vital for managing airport resources and services efficiently. | * Window * CTEs * Aggregations * Advanced Joins |
| Weekday Flight Frequency Analysis | To determine the average number of flights operating from a specific source to a destination airport on a given weekday | This analysis aims to provide insights into the frequency of flights on popular routes during weekdays, which can be crucial for optimizing flight schedules and understanding market demand | * Time functions * Advanced Joins * Aggregations * CTEs |
| Popularity Ranking of Airlines by Destination | To rank airlines based on their popularity for specific destinations | This analysis aims to identify which airlines dominate routes, providing insights into market shares and consumer preferences | * Windows * CTEs * Advanced Joins |
| Airline Ranking Based on Average Flight Frequency from a Specific Source Airport | To rank airlines based on the average number of flights they operate from a given source airport | This analysis aims to identify which airlines are most active at a particular airport, providing insights into their operational focus and market presence at that location. | * Partition * Window * Aggregations |
| Flight Duration and Pricing Correlation Analysis | To examine the relationship between flight duration and ticket pricing across different airlines | This analysis aims to understand how flight length impacts fare structures, a key component in developing competitive pricing strategies | * CTEs * Aggregations |
| Analysis of Traffic by Airport Category and Size | To evaluate the scalability of airports by analyzing passenger counts for domestic and international flights in relation to airport size. | This analysis will help in understanding the capacity and traffic distribution across different sizes of airports, typically with international flights associated with larger airports and domestic flights with smaller ones. | * Regexp * Case when * Aggregations |

**7. Conclusion**

Concluding our comprehensive analysis for the strategic entry of a new airline in the Indian aviation market, we have successfully delivered key insights aligned with our project objective. By meticulously examining data on average economy pricing, flight frequencies, airport traffic, departure times, flight duration-pricing relationships, and airline popularity by destination, we have provided a detailed, data-driven foundation for our client's market entry strategy. These deliverables, grounded in advanced SQL analytics, offer a holistic view of the competitive landscape and consumer behavior, enabling our client to make informed decisions on pricing, route selection, and operational planning. This strategic approach positions the new airline not only to effectively compete with existing market challengers but also to lay a robust groundwork for future scaling to challenge the market leaders, ensuring a strong and sustainable presence in the dynamic Indian aviation sector.

**8. Appendix**

# Query 1 Flight Information by Code and Source

DELIMITER &&

CREATE PROCEDURE GetAvgEconomyPrice (IN Flight\_Code VARCHAR(10), IN Source\_ID INT)

BEGIN

SELECT a.Airline AS Airline\_Name, ap.Airport\_name AS Source\_Airport\_Name, ap.Place AS Source\_Place\_Name,

AVG(ti.Economy) AS Average\_Economy\_Price

FROM travel\_info ti

INNER JOIN airlines a

ON ti.Flight\_code = a.Flight\_code

INNER JOIN airports ap

ON ti.Source\_ID = ap.Airport\_ID

WHERE ti.Flight\_code = Flight\_Code AND ti.Source\_ID = Source\_ID

GROUP BY a.Airline, ap.Airport\_name, ap.Place;

END &&

DELIMITER ;

CALL GetAvgEconomyPrice('9I-894', 26618);

# Query 2 Flight Information by Source and Destination IDs

DELIMITER &&

CREATE PROCEDURE GetAirlinesWithPlacesAveragePrice(IN Source\_ID INT, IN Destination\_ID INT)

BEGIN SELECT a.Airline AS Airline\_Name, src\_ap.Airport\_name AS Source\_Airport\_Name, src\_ap.Place AS Source\_Place\_Name, dest\_ap.Airport\_name AS Destination\_Airport\_Name, dest\_ap.Place AS Destination\_Place\_Name, AVG(ti.Economy) AS Average\_Economy\_Price

FROM travel\_info ti

INNER JOIN airlines a ON ti.Flight\_code = a.Flight\_code

INNER JOIN airports src\_ap ON ti.Source\_ID = src\_ap.Airport\_ID

INNER JOIN airports dest\_ap ON ti.Destination\_ID = dest\_ap.Airport\_ID

WHERE ti.Source\_ID = Source\_ID AND ti.Destination\_ID = Destination\_ID

GROUP BY a.Airline, src\_ap.Airport\_name, src\_ap.Place, dest\_ap.Airport\_name, dest\_ap.Place;

END && DELIMITER ;

CALL GetAirlinesWithPlacesAveragePrice('35145', '35141');

# Query 3 Top 5 Destinations from Source Airport

DELIMITER &&

CREATE PROCEDURE TopDestinationFromSource(IN SourceAirportID INT)

BEGIN

SELECT

dest\_ap.Airport\_name AS Destination\_Airport\_Name, dest\_ap.Place AS Destination\_place,

COUNT(ti.Flight\_code) AS NumberOfFlights

FROM travel\_info ti

INNER JOIN airports dest\_ap ON ti.Destination\_ID = dest\_ap.Airport\_ID

WHERE ti.Source\_ID = SourceAirportID

GROUP BY ti.Destination\_ID, dest\_ap.Airport\_name, dest\_ap.Place

ORDER BY NumberOfFlights DESC

LIMIT 5;

END &&

DELIMITER ;

CALL TopDestinationFromSource('26434');

# Query 4 Airport Rank Hierarchical Structure

WITH RECURSIVE AirportHierarchy AS (

SELECT Airport\_ID, Airport\_name, Place, Ranking

FROM airports

WHERE Ranking = (SELECT MIN(Ranking) FROM airports)

UNION ALL

SELECT a.Airport\_ID, a.Airport\_name, a.Place, a.Ranking

FROM airports a

INNER JOIN AirportHierarchy ah ON a.Ranking = ah.Ranking + 1)

SELECT \* FROM AirportHierarchy limit 8;

# Query 5 Top Airports Ranked by Male and Female Literates

SELECT a.Airport\_ID, a.Airport\_name, a.Place, a.ranking,

(SELECT b.literates\_male FROM place b WHERE a.Place = b.city) AS literates\_male,

(SELECT b.literates\_female FROM place b WHERE a.Place = b.city) AS literates\_female

FROM airports a

ORDER BY literates\_male DESC, literates\_female DESC

LIMIT 10;

# Query 6 Top 10 Airports by Traffic & Gender Population

SELECT

a.Airport\_ID, a.Airport\_name, a.Place, a.Passengers\_2022\_23 as passengertraffic\_2022\_23,

(SELECT b.population\_male FROM place b WHERE a.Place = b.city) AS population\_male,

(SELECT b.population\_female FROM place b WHERE a.Place = b.city) AS population\_female

FROM airports a

ORDER BY population\_male DESC, population\_female DESC

LIMIT 10;

# Query 7 Airports and their busiest departure times

WITH DepartureArrivalTimes AS (

SELECT a.Airport\_name, EXTRACT(HOUR FROM t.Departure\_Time) AS DepartureHour,

COUNT(\*) OVER (PARTITION BY a.Airport\_ID, EXTRACT(HOUR FROM t.Departure\_Time)) AS DepartureCount,

COUNT(\*) OVER (PARTITION BY a.Airport\_ID, EXTRACT(HOUR FROM t.Arrival\_Time)) AS ArrivalCount

FROM travel\_info t

JOIN Airports a ON t.Source\_ID = a.Airport\_ID OR t.Destination\_ID = a.Airport\_ID

),

RankedDepartureTimes AS (

SELECT Airport\_name, DepartureHour, DepartureCount,

RANK() OVER (PARTITION BY Airport\_name ORDER BY DepartureCount DESC) AS DepartureRank

FROM DepartureArrivalTimes

)

SELECT DISTINCT d.Airport\_name, CONCAT(d.DepartureHour,":00") AS BestDepartureHour

FROM RankedDepartureTimes d

WHERE d.DepartureRank = 1

ORDER BY d.Airport\_name;

# Query 8 Average #Flights Source to Destination on Weekday

WITH DailyFlights AS (

SELECT DAYNAME(STR\_TO\_DATE(t.Date\_of\_journey, '%d-%m-%Y')) AS Weekday, s.Airport\_name AS SourceAirport,

d.Airport\_name AS DestinationAirport, COUNT(\*) AS FlightsOnDay

FROM travel\_info t

JOIN airports s ON t.Source\_ID = s.Airport\_ID

JOIN airports d ON t.Destination\_ID = d.Airport\_ID

GROUP BY Weekday, SourceAirport, DestinationAirport

),

RankedFlights AS ( SELECT Weekday, SourceAirport, DestinationAirport, AVG(FlightsOnDay) AS AvgFlights,

RANK() OVER (PARTITION BY Weekday ORDER BY AVG(FlightsOnDay) DESC) AS Day\_Rank

FROM DailyFlights

GROUP BY Weekday, SourceAirport, DestinationAirport

)

SELECT Weekday, SourceAirport, DestinationAirport, AvgFlights

FROM RankedFlights

WHERE Day\_Rank <= 3 -- Filter to show only top 3 ranks per weekday

ORDER BY Weekday, Day\_Rank;

# Query 9 Ranking Airlines by Popularity per Destination

WITH FlightCounts AS (

SELECT a.Airline AS airline, ti.destination\_id AS destination, ap.Place, COUNT(\*) AS flights\_count

FROM travel\_Info ti

JOIN airlines a ON ti.flight\_code = a.flight\_code

JOIN airports ap ON ap.airport\_ID = ti.destination\_id

GROUP BY a.Airline, ti.destination\_id, ap.Place),

RankedAirlines AS

(SELECT airline, destination, Place, flights\_count,

RANK() OVER(PARTITION BY destination ORDER BY flights\_count DESC) AS rank\_destination

FROM FlightCounts)

SELECT destination, Place, airline, flights\_count, rank\_destination

FROM RankedAirlines

WHERE rank\_destination <= 3;

# Query 10 Airline Ranking by Scheduled Trips per Airport

WITH AirlineFlightCounts AS (

SELECT ti.Source\_ID, a.Airline, COUNT(ti.Flight\_code) AS FlightCount,

AVG(ti.Economy) AS AvgEconomyPrice

FROM travel\_info ti

JOIN airlines a ON ti.Flight\_code = a.Flight\_code

GROUP BY ti.Source\_ID, a.Airline

),

RankedAirlines AS (

-- Rank airlines by flight count for each airport

SELECT Source\_ID, Airline, FlightCount, AvgEconomyPrice,

RANK() OVER (PARTITION BY Source\_ID ORDER BY FlightCount DESC) AS rank\_airline

FROM AirlineFlightCounts

)

SELECT ra.Source\_ID, ap.Airport\_name AS Airport\_Name, ra.Airline, ra.FlightCount, ra.AvgEconomyPrice, ra.rank\_airline

FROM RankedAirlines ra

JOIN airports ap ON ra.Source\_ID = ap.Airport\_ID

ORDER BY ra.Source\_ID, ra.rank\_airline;

# Query 11 Analyzing Flight Duration and Pricing Relationship

CREATE VIEW DurationPricing AS

SELECT ti.destination\_ID, AVG(ti.economy) AS average\_price, AVG(duration\_in\_mins) AS average\_duration

FROM travel\_Info ti

GROUP BY ti.destination\_ID; -- Query the view

SELECT dp.destination\_ID AS Destination\_ID, ap.place AS Destination, dp.average\_price, dp.average\_duration

FROM DurationPricing dp

JOIN airports ap ON ap.airport\_ID = dp.destination\_id

ORDER BY average\_duration DESC;

# Query 12 Analysis of Traffic by Airport Category and Size

SELECT

CASE

WHEN Airport\_name REGEXP '[Ii]nternational' THEN 'International'

ELSE 'Domestic'

END AS Airport\_Category,

type AS Size\_Category,

SUM(Passengers\_2022\_23) AS Total\_Passengers\_2022\_23,

SUM(Passengers\_2021\_22) AS Total\_Passengers\_2021\_22,

ROUND(

((SUM(Passengers\_2022\_23) - SUM(Passengers\_2021\_22)) / SUM(Passengers\_2021\_22)) \* 100, 2

) AS Percentage\_Growth

FROM airports

GROUP BY Airport\_Category, Size\_Category

ORDER BY Airport\_Category, Size\_Category;